

Remarks and Arguments

Claims 1-30 have been presented for examination.

Claims 1-30 have been rejected under 35. U.S.C. §102(e) as anticipated by U.S. Patent No. 6,535,871 (Romansky.) The examiner comments that all of the recited elements are disclosed in the Romansky reference.

The present invention relates to a method and apparatus that allows the contents of secure documents to be indexed by conventional search engines without making the plain text available. In particular, a text stream derived from the entire document content is broken into two to five word phrases, the phrases are then randomized and a text file is created from the randomized stream. This process produces a scrambled text file that contains nearly all of the words in the original document and most of the phrases, but the scrambled file cannot be read by humans. In particular, the word frequency and word context is largely retained. Third party search engines are allowed to index the scrambled file so that search algorithms that search on particular words or phrases produce nearly the same number of hits as with the plain text file.

The Romansky reference discloses a method and apparatus for generating a plain text index of a secure document. As disclosed, the index is generated by extracting keywords from the plain text of the secure document, eliminating “problematic” keywords, such as names, from the resulting keyword list to produce a reduced index and then scrambling the keywords in the reduced index. This is set forth in Romansky, column 2, line 66 – column 3, line 34 in relation to Figure 1. Since the Romansky index generating method uses single keywords instead of multi-word phrases, the method loses most of the word context. This is noted in Romansky at column 1, line 66 – column 2, line 1. Thus, search algorithms that search on particular words or phrases will generally not produce nearly the same number of hits as with the plain text file with the Romansky scrambled word index.

These differences are recited in the claims. Claim 1 is illustrative. It recites, in lines 5-6, “...fragmenting the text stream into multi-word phrases” and “randomly assembling the phrases into a scrambled document...” As discussed above, Romansky generates a keyword list and then scrambles the words on that list. The examiner points to Romansky, column 2, lines 11-32 and column 3, lines 6-17 as disclosing

fragmenting a text stream into multi-word phrases. However, at these locations, Romansky is clearly discussing words “out of context” or “tokens” (words with leading and trailing spaces removed, and with punctuation and duplicates removed). There is no discussion of using multi-word phrases. Similarly, Romansky scrambles words not phrases. The examiner points to Romansky, column 2, lines 33-40 and column 3, lines 24-35 as disclosing this step. However, at these locations it is clear that Romansky mentions phrases in connection with methods for insuring that they do not occur in the output in order to prevent information that might be contained in the phrases from appearing in the scrambled index. Romansky column 3, lines 24-26, states that the tokens are randomized. Thus, claim 1 recites steps not disclosed in Romansky and thereby distinguishes over the Romansky reference.

Claims 2-10 are dependent on claim 1 and include the recited steps. Therefore, they also distinguish over the cited Romansky reference in same manner as claim 1. In addition, these claims recite additional limitations not disclosed in the Romansky reference. For example, claims 2-4 recite that the multi-word phrases contain, at least two words, a random number of words and a maximum of five words, respectively. In Romansky all word tokens consist of a single word. Thus, claims 2-4 distinguish over the cited reference for this reason also. In addition, claim 5 recites that the position of word phrases is swapped in the text stream. In Romansky, word positions are swapped. Thus, claim 5 also distinguishes over the cited reference in the same manner as claim 1.

Claim 11 is an apparatus claim that contains limitations that parallel those in claim 1. Therefore, it distinguishes over the cited Romansky reference in the same manner as claim 1. Claims 12-20 are dependent on claim 11 and include the recited limitations. Therefore, they also distinguish over the cited Romansky reference in same manner as claim 11. In addition, these claims recite additional limitations not disclosed in the Romansky reference. For example, claims 12-15 recite limitations that parallel those recited in claims 2-5 and distinguish over the cited reference in a manner similar to that discussed above.

Claim 21 is a computer program product claim that contains limitations that parallel those in claims 1 and 11. Therefore, it distinguishes over the cited Romansky

reference in the same manner as claims 1 and 11. Claims 22-30 are dependent on claim 21 and include the recited limitations. Therefore, they also distinguish over the cited Romansky reference in same manner as claim 21. In addition, these claims recite additional limitations not disclosed in the Romansky reference. For example, claims 22-25 recite limitations that parallel those recited in claims 2-5 and distinguish over the cited reference in a manner similar to that discussed above.

In light of the forgoing amendments and remarks, this application is now believed in condition for allowance and a notice of allowance is earnestly solicited. If the examiner has any further questions regarding this amendment, she is invited to call applicants' attorney at the number listed below. The examiner is hereby authorized to charge any fees or direct any payment under 37 C.F.R. §§1.17, 1.16 to Deposit Account number 50-3969.

Respectfully submitted

/paul e. kudirka/ Date: 2007-01-04
Paul E. Kudirka, Esq. Reg. No. 26,931
LAW OFFICES OF PAUL E. KUDIRKA
Customer Number 64967
Tel: (617) 357-0010 Fax: (617) 357-0035